

“Ok Google, am I sick?”: artificial intelligence, e-health, and data protection regulation

Paolo Guarda*

ABSTRACT: The healthcare sector has been impacted by Artificial Intelligence systems for several years. This is certainly favoured by the constant development of technologies based on sophisticated machine learning and techniques that are able to find complex patterns in data. However, several critical issues hinder AI’s definitive affirmation: from a lack of clinical studies that can demonstrate its reliability and greater effectiveness with respect to traditional systems, to the criticalities related to the attribution of responsibility in the event of medical errors. In particular the application of data protection regulation to this specific scenario needs to be carefully evaluated. The correct management of this technology, through legal intervention in preventing possible dehumanization, plays a crucial role. This paper aims to investigate the impact of technologies based on AI in the healthcare sector, with particular attention to personal data protection issues.

KEYWORDS: artificial intelligence; e-Health; privacy; data protection; automated decision making.

SUMMARY: 1. Artificial Intelligence: science fiction or reality? – 2. Artificial Intelligence in the e-health sector – 3. Data protection regulation issues – 3.1 Prohibition of automated decision – 3.2 A right to explanation or only a right to be informed? – 4. Final remarks: algorithms, empathy and the importance of making mistakes.

1. Artificial Intelligence: science fiction or reality

The topic of Artificial Intelligence (AI) lies somewhere between literary and cinematic fiction and reality. For decades, novelists and film directors have told us about science fiction worlds where machines with a form of evolved intelligence interact with humans. In science fiction, AI is a recurring theme, as a simple narrative element or a central topic of history. It is generally presented in the form of advanced computers, robots or androids. To mention just a few of the countless examples: the masterpiece *AI Artificial Intelligence*, a film directed by Steven Spielberg in 2001 and based on a project by Stanley Kubrick, which tells the story of a “Mecha”, an automaton produced by a major company, with the features and behaviour of a child that is able to experience a true feeling of love for a human being; or *Ex Machina*, a film written and directed by Alex Garland in 2015, about a humanoid with feminine features endowed with true intelligence and self-awareness. Among sentient computers we may recall, for example, the “*Multivac*”, featured in some stories by

* Assistant Professor of Private Comparative Law. Mail: paolo.guarda@unitn.it. The article was subject to a double-blind peer review process.

Isaac Asimov, comparable to modern grid computing systems, and "HAL 9000" from the film *2001: A Space Odyssey* (1968) by Stanley Kubrick.

Most of the narrated scenarios turn out to be, in the end, dystopias describing forms of unintended interactions, attempts at the prevarication of machines against humans, the failure of dreams of a bright and futuristic progress. The theme is, then, often linked to the classic rebellion of the machine, in which a computer (in most cases sentient) turns against the human beings who had built it. The question that may arise now is: why do we start a paper dealing with the use of AI in the digital health sector in this way? Now, the reason is linked to the fact that, beyond the necessary literary and cinematographic exaggerations, these stories always tell what may happen when the impact of new technologies in the human context is not carefully and thoughtfully governed. The use of artificial intelligence in the health sector is no longer a mirage or the description of a visionary scenario. It is becoming more and more daily practice and is expressed in: platforms that are able to provide answers and advice increasingly supporting doctors in the treatment of particular types of diseases; telemedicine apps empowering "virtual coaching" systems that tend to replace the traditional medical-patient interaction; devices that interact as chatbots with patients and support them in various ways, providing monitoring activities of their health status too, and so on.

What do we mean by the term "Artificial Intelligence"? Not all AI is created equal, to quote a fascinating version, widely used in online contributions dealing with the issue, of one of the most famous American mottos.¹ There are, indeed, many versions, definitions and concepts of AI. Purists and technicians of the field debate on what can truly be understood as "intelligent" and what is not at all, or on what is the correct terminology to refer to this type of technology.²

For example, some virtual assistants, such as Siri or Google Assistant, are often referred to as "AI", but they may, actually, be classified as "weak" or "narrow", since they are basically designed to organize information and answer questions in a relatively limited context. In fact, these assistants implement a combination of speech recognition, natural language processing and AI to perform actions such as using voice commands to retrieve maps, search for events in the calendar and find information.³ On the other hand, other forms of AI are simply beyond current technical possibilities and appear as possible, future scenarios of application. In order to reach human levels of reasoning, AI should be equipped with recursive self-improvement and capable of identifying and solving complex

¹ The first use of this motto is traditionally traced back to Thomas Jefferson in the United States Declaration of Independence and was penned in 1776 at the beginning of the American Revolution.

² See, among others, S. LEGG, M. HUTTER, *A Collection of Definitions of Intelligence (Technical report)*, IDSIA, 2007, available at: [arXiv:0706.3639](https://arxiv.org/abs/0706.3639) (last visited 10 November 2018); S.J. RUSSELL, P. NORVIG, *Artificial Intelligence: A Modern Approach*, (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, 2003 (who prefer the term "rational agent" and write "*The whole-agent view is now widely accepted in the field*", p. 55); D. POOLE, A. MACKWORT, R.R. GOEBEL, *Computational Intelligence: A Logical Approach*, New York: Oxford University Press, 1998 (where the term "computational intelligence" is used as a synonym of AI); N. NILSSON, *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann, 1998.

³ See J. CHUNG, *What Should We Do About Artificial Intelligence in Health Care?*, in *NYSBA Health Law Journal*, vol. 22, No. 3, 2017, 37 (available at: <https://ssrn.com/abstract=3113655>, last visited 10 November 2018).

problems on its own, without a need for human intervention. According to some commentators, this type of technology, properly called “strong” AI, has a long way to come.⁴

Without any claim to resolving the terminological dispute, we may refer to the definition proposed by the Communication of the European Commission on “Artificial Intelligence for Europe”: “*Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals*”.⁵ The term AI is, in general, applied when a machine mimics “cognitive” functions that humans associate with other human minds, such as “learning” and “problem solving”. In order to emphasize the aspects that will become relevant later on in the legal analysis, we can state that the main characteristics are: the collection of large amounts of information (this aspect evidently may be referred to as so-called Big Data) and the ability to take autonomous decisions and actions aimed at maximizing the chances of success.

This paper aims to provide an initial investigation of the impact of AI-based technologies on the healthcare sector, with particular attention to critical data protection issues and in full awareness that it will not always be possible to give reliable and conclusive answers to the questions raised. Thus, the second section will be dedicated to presenting the possible applications in the field of digital health, pinpointing the current trends, the expected or promised benefits and the aspects to be evaluated carefully. The third section will go into the details of the legal analysis highlighting the impact of AI in terms of personal data protection regulation. One of the major concerns of these instruments, the so called “black boxes”, will be highlighted: it refers to the use of algorithms, whose logics are generally hidden from users and which result in a total lack of transparency about their operational logics. Finally, the conclusions will pull the strings of the essay by focusing on cultural and social aspects of this phenomenon.

2. Artificial Intelligence in the e-health sector

The healthcare sector has been impacted by AI technologies for several years. This is certainly favoured by the constant development of technologies based on sophisticated machine-learning and AI techniques that are increasingly able to find complex patterns in data. Another phenomenon has also contributed to the explosion of this type of approach: the enormous availability of information that the era of Big Data allows.⁶ The possible sources of this information are the most disparate: Elec-

⁴ See D. GALEON, C. REEDY, *Kurzweil Claims That the Singularity Will Happen by 2045*, *Futurism*, October 5, 2017, available at <https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045> (last visited 10 November 2018).

⁵ Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, *Artificial Intelligence for Europe*, Brussels, 25 April 2018 COM (2018) 237 final.

⁶ The definition of Big Data is the subject of intense debates among scholars. The Big Data phenomenon is characterized by the so-called “three Vs”: “Volume” (large quantity of data), “Variety” (heterogeneity in the data), and “Velocity” (fast access to the data). Another two “Vs” have been added: “Veracity” (the messiness or trustworthiness of the data) and “Value” (the economic value that data may have). For a definition of Big Data, among others, see A. DE MAURO, M. GRECO, M. GRIMALDI, *A Formal definition of Big Data based on its essential features*, in *Library Review*, vol. 65, No. 3, 2016, pp. 122-135; B. MARR, *Big Data: Using Smart Big Data, Analytics*

tronic Health Records (EHRs), medical literature, clinical trials, insurance claims data, pharmacy records, data entered by patients or recorded on fitness trackers, etc.⁷ The confluence of these two factors has caused a considerable push towards the development and refinement of personalized medicine methods.

There are many expectations and hopes for applying AI techniques in the medical field. First of all, the possibility of developing predictive models with obvious advantages in terms of prevention. Secondly, the ability to put in place early diagnosis, in order to guarantee a prompt reaction using the most appropriate care. Finally, the affirmation of chatbot-based environments promises to guarantee the right information for patients, accompanying them in their care processes.

AI could be capable of solving many problems in several e-health contexts.

Intelligent virtual assistants, now embedded in smartphones or dedicated home speakers, like Microsoft Cortana and Apple Siri, but increasingly present in people's homes devices, such as voice assistants Amazon Alexa or Google Assistant, are supported by systems with powerful AI features. These tools represent today the most advanced and captivating frontier of the use of AI to facilitate everyday life. The title of this contribution was inspired by this technology which, to tell the truth, has not yet seen such a massive adoption. Voice assistants are software agents that can interpret human speech and respond via synthesized voices. Users are able to ask questions, control devices and media playback via voice, and manage other basic tasks such as email and calendars with verbal commands. There are of course several privacy and security issues inherent to voice assistants as potential future uses for these devices. They can, indeed, have an added value, especially when combined with health care apps, ensuring a greater integration of AI tools into daily medical applications, thus improving treatment efficiency, avoiding costs and minimizing the risks of false diagnosis. All of this can also evidently facilitate a more targeted pre-operative planning or support the post-operative phases.⁸

and Metrics to Make Better Decisions and Improve Performance, John Wiley & Sons, Ltd.; 1st edition (1 Feb. 2015). For further details, especially with reference to the legal issues, see, among others, B. RABAI, *I "big data" nell'ecosistema digitale: tra libertà economiche e tutela dei diritti fondamentali*, in *Amministrare*, 2017, vol. 3, 405-422; M. BOGNI, A. DEFANT, *Big Data: diritti IP e problemi della privacy*, in *Dir. industriale*, 2015, fasc. 2, 117; S. HOFFMAN, A. PODGURSKI, *Big Bad Data: Law, Public Health, and Biomedical Databases*, in *J Law Med Ethics*, 2013, vol. 41, Suppl 1, 56-60; J. POLONETSKY, O. TENE, *Privacy and Big Data: Making Ends Meet*, 66 *Stan. L. Rev.* 25 (2013).

⁷ See W. NICHOLSON PRICE, *Artificial Intelligence in Health Care: Applications and Legal Implications*, in *The SciTech Lawyer*, vol. 14, no. 1, 2017, p. 10.

⁸ With reference to the use of voice assistants such as Alexa, Siri and Cortana in the medical context and with particular attention to the connected privacy and security issues, see A. PFEIFLE, *Alexa, What Should We Do about Privacy? Protecting Privacy for Users of Voice-activated Devices*, 93 *Wash. L. Rev.* 421 (2018); M. E. STUCKE, A. EZRACHI, *Alexa et al., What Are You Doing with My Data?*, in *Critical Analysis of Law*, vol. 5, no. 1, 2018, 148; M.B. HOY, *Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants*, in *Medical Reference Services Quarterly*, vol. 37, no. 1, 2018, 81; A. HASSOON, J. SCHRACK, D. NAIMAN, D. LANSEY, Y. BAIG, V. STEARNS, D. CELENTANO, S. MARTIN, L. APPEL, *Increasing Physical Activity Amongst Overweight and Obese Cancer Survivors Using an Alexa-Based Intelligent Agent for Patient Coaching: Protocol for the Physical Activity by Technology Help (PATH) Trial*, in *JMIR Res Protoc*, vol. 7, iss. 2, e27, 2018, p. 2, available at: <https://www.researchprotocols.org/2018/2/e27/> (last visited 10 November 2018).

Particular conditions requiring elaborate treatment plans could benefit from AI tools during specific therapies. Incorporating an artificial intelligence system capable of automatically formulating plans based on specific conditions would provide value to both physicians and patients alike. This is the case of ecosystems designed to innovate the processes of interaction between physician and patient, with an evident impact also on the organizational models that regulate the performance of health services.⁹ These intelligent systems may look for inconsistencies, errors and omissions in an existing treatment plan or may be used to formulate a treatment based on a patient's specific condition and on accepted treatment guidelines.

The intelligent agent can, furthermore, be used for finding information, for example on the Internet, relevant to a particular disease, integrating knowledge about user preferences and needs into such searches, or for automatically interpreting medical images (i.e. X-rays, angiograms, CT scans, etc.), becoming a formidable tool especially with regard to mass screens where the system could signal potentially abnormal images for detailed human attention.

Finally, expert systems and decision support systems, programmed to aggregate and store a large amount of data modelled for specific purposes, can be successfully used in the field of medical devices in applications for cardiac monitoring and automated ECG, medical imaging, clinical laboratory analysis, and so on.

A well-known tool, even though we may debate its real effectiveness, in the medical industry is the famous “IBM Watson”,¹⁰ which represents a sort of paradigmatic example of AI implemented in this sector. Watson is a point between weak and strong systems. Commonly referred to as an “expert system,” it uses AI to solve defined problems in a specialized subject area. These expert tools can leverage “fuzzy logic” to go beyond binary “yes/no” and “true/false” questions to tackle tasks such as logic games, financial investing, legal research, and medical research. Watson endorses “DeepQA” and has a software architecture that analyses, reasons about, and answers to the content fed into Watson. It can be fed by information on oncological matters – namely 300 medical journals, 200 textbooks and nearly 15 million pages of text – and scour them to present doctors with treatment options, and recommended drugs and instructions for administration. This raw capability of parsing through data is combined with “training” by doctors at institutions.¹¹

⁹ A tool of this type is represented, for example, by the “TreC_Diabete” platform, promoted by the Department of Health and Social Policies of the Autonomous Province of Trento in collaboration with the Provincial Health Care Provider (APSS) and realized under the technical-scientific management of the Bruno Kessler Foundation (FBK). “TreC_Diabete” is a system for supporting self-care and remote monitoring of patients with diabetic pathology. It features a technological platform that allows, on the one hand, to interoperate with third-party systems (eg. EHR, Medical Health Records, Hospital Information System, cloud repository, etc.) and, on the other hand, to activate applications addressed to physicians and patients. “TreC_Diabete” includes a web application aimed at health professionals (clinical dashboard) and a mobile application (for Android and iOS) for the patient (patient APP). For further description see D. CONFORTI, “Le prescrivo una app”: come in trentino curano i malati cronici, February 2017, available at: <https://www.agendadigitale.eu/infrastrutture/le-prescrivo-un-app-ecco-come-in-trentino-curano-i-malati-cronici/> (last visited 10 November 2018).

¹⁰ <https://www.ibm.com/watson/>.

¹¹ See J. CHUNG, *What Should We Do About Artificial Intelligence in Health Care?*, op. cit., pp. 37-38. The author also says that: “Watson, then, is far from perfect. (...) the fallibility of the people both tasked with initially programming Watson and the doctors training Watson. And some errors may simply occur randomly due to the unpredictable nature of how machines process and organize unfamiliar data” (p. 38). For further details on IBM

The premises for more extensive AI uses in health care are all there. However, some critical issues hinder its definitive affirmation. There is a lack of clinical studies that can demonstrate reliability and greater effectiveness than traditional systems in making predictions, diagnoses or suggesting appropriate therapies: this causes a certain mistrust on the part of the physicians towards their use and a limit also for policy makers in pushing for their definitive adoption.¹² Moreover, the issue related to the attribution of responsibility in the event of medical errors still presents critical aspects that have not yet been fully analysed and resolved by the legal scholarship (the intervention of lawmakers could likely provide clarity).¹³ Then there are the issues related to the protection of privacy and security on which this writing will focus.¹⁴ Finally, the ethical aspects play an essential role, not only in

Watson see J. CHUNG, A. ZINK, *Hey Watson, Can I Sue You for Malpractice? Examining the Liability of Artificial Intelligence in Medicine*, in *Asia-Pacific Journal of Health Law, Policy and Ethics*, 2017, forthcoming (available at <https://ssrn.com/abstract=3076576>, last visited 10 November 2018); W. SHIH, *Building Watson: It's Not So Elementary My Dear!*, Harvard Business School Technology & Operations Mgt. Unit Case No. 612-017, 2011, available at: <https://ssrn.com/abstract=1993074> (last visited 10 November 2018); B. COOPER, *Judges in Jeopardy! Could IBM's Watson Beat Courts at Their Own Game?*, 121 *Yale L.J. Online* 87 (2011).

¹² See E. SANTORO, *Intelligenza artificiale in Sanità, ecco tutti i nodi da sciogliere*, in *AgendaDigitale*, 2017, available at <https://www.agendadigitale.eu/sanita/intelligenza-artificiale-in-sanita-ecco-tutti-i-nodi-da-sciogliere/> (last visited 10 November 2018).

¹³ Currently there is no clear legal or regulatory regime with regards to liability for AI. Newspapers and TV reported that in June 2016 a man died while driving in autopilot mode because the sensors of the vehicle, which help to steer the car by identifying obstructions, had failed to recognize “the white side of the tractor trailer against a brightly lit sky” (see THE TESLA TEAM, *A tragic loss*, 31 June 2016, available at <https://www.teslamotors.com/blog/tragic-loss> (last visited 10 November 2018)). The US National Highway Traffic Safety Administration then opened a preliminary evaluation into the performance of the autopilot, so it is too soon to reach any conclusions. The theories of recovery under tort were usually divided by personhood: in the common law system, for instance, negligence-based tort law, such as medical malpractice or vicarious liability, generally applies to actions of persons owing a duty of care; on the other hand, the regime of product liability applies to things and generally holds the manufacturer strictly liable for defective products. Furthermore, according to US courts, robots cannot be sued: see *United States v. Athlove Industries, Inc.*, 746 F.2d 977 (3rd Cir. 1984). For further details see J. CHUNG, *What Should We Do About Artificial Intelligence in Health Care?*, *op. cit.*, 38-39. With reference, for example, to a civil law context such as the Italian legal system, the main liability regimes that apply to machines are breach of contract (*responsabilità contrattuale*), torts (*responsabilità extracontrattuale*) and product liability. The applicative difficulties remain the same however: See G. NOTO LA DIEGA, *Machine Rules. Of Drones, Robots, and the Info-Capitalist Society*, in *The Italian Law Journal*, Vol. 2, No. 2, 2016, 367, 394-400. In general with reference to liability for AI, see G. WAGNER, *Robot Liability*, June 19, 2018, available at <https://ssrn.com/abstract=3198764> (last visited 10 November 2018); N. PETIT, *Law and Regulation of Artificial Intelligence and Robots - Conceptual Framework and Normative Implications*, March 9, 2017, available at <https://ssrn.com/abstract=2931339> (last visited 10 November 2018); European Parliament Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), with particular attention to paragraph 33 dedicated to “Medical Robots”; K. SHERIFF, *Defining Autonomy in the Context of Tort Liability: Is Machine Learning Indicative of Robotic Responsibility?*, December 12, 2015, available at <https://ssrn.com/abstract=2735945> (last visited 10 November 2018). With specific reference to the issues related to the liability that derives from the use of telemedicine tools, see C. BOTRUGNO, *Un diritto per la telemedicina: analisi di un complesso normativo in formazione*, in *Politica del diritto*, No. 4, 2014, 639, 658-663. Finally, some legal scholars have already detected the difficulties of adapting the current legal and regulatory system to AI and proposed that a new legal regime be created: see, among others, J.S. ALLAIN, *From Jeopardy! To Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems*, 73 *La. L. Rev.* 100 (2013).

¹⁴ See below section 3.



terms of job losses (or rather of transformation of these), but especially for the future, and in part dystopian, scenarios that the emergence of these technologies sometimes suggest.¹⁵

We may quote some general statements or white papers on AI. Among these, the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe, 25 April 2018 (COM/2018/237 final) and the recent “Universal Guidelines on Artificial Intelligence (UGAI)”, 23 October 2018 Brussels, Belgium,¹⁶ which call attention to the growing challenges of intelligent computational systems and proposals that can improve and inform their design. In its core, the purpose of the UGAI is to promote transparency and accountability for these systems. Not all systems fall within the scope of these Guidelines. The concern is with those systems that affect the rights of people. Above all else, these systems should do no harm. Furthermore, it is worth mentioning the “Declaration on Ethics and Data Protection in Artificial Intelligence” from the “2018 International Conference of Data Protection and Privacy Commissioners”, adopted in Brussels on the same day, endorsing some guiding principles as its core values in preserving human rights in the development of artificial intelligence.¹⁷

AI in the health sector is an issue that concerns and primarily involves the public stakeholder, whose task is to act as an attractor of technological innovation based on the appropriate tools to complete paths of total digitalization still underway and unfinished for too long, but that are prodromal to any further path of innovation. In this regard, for example in the Italian context, the first “White Paper - Artificial Intelligence at the service of the citizen” was presented on 21 March 2018 in Rome by the AI task force of the Italian Digital Agency. The White Paper illustrates the guidelines and recommendations for the sustainable and responsible use of Artificial Intelligence in Public Administration and represents the beginning of a project path¹⁸. The White Paper identifies nine “Challenges”: Ethics, Technology, Skills, Data Roles, Legal Context, Accompanying Transformation, Preventing Inequalities, Measuring Impact, Human Being. It also draws up a handbook of recommendations to trace the next

¹⁵ As regards ethical issues in AI, see S. BAUM, *Social Choice Ethics in Artificial Intelligence*, in *AI & Society*, forthcoming, available at <https://ssrn.com/abstract=3046725> (last visited 10 November 2018); B. MITTELSTADT, P. ALLO, M. TADDEO, S. WACHTER, L. FLORIDI, *The Ethics of Algorithms: Mapping the Debate*, in *Big Data & Society*, Vol. 3(2), 2016, available at <https://ssrn.com/abstract=2909885> (last visited 10 November 2018); S. BIRD, S. BAROCAS, K. CRAWFORD, F. DIAZ, H. WALLACH, *Exploring or Exploiting? Social and Ethical Implications of Autonomous Experimentation in AI*, Workshop on Fairness, Accountability, and Transparency in Machine Learning, 2016, available at <https://ssrn.com/abstract=2846909> (last visited 10 November 2018); D. HELBING, *Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies*, April 14, 2015, available at SSRN: <https://ssrn.com/abstract=2594352> (last visited 10 November 2018); P. LIN ET AL. (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics*, Cambridge, MIT Press, 2011; C. ALLEN ET AL., *Prolegomena to Any Future Artificial Moral Agent*, in *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 12, 2000, p. 251.

¹⁶ <https://thepublicvoice.org/ai-universal-guidelines/> (last visited 10 November 2018).

¹⁷ https://edps.europa.eu/sites/edp/files/publication/icdppc-40th_ai-declaration_adopted_en_0.pdf (last visited 26 November 2018). The declaration will be open on public consultation. See also by the High Level Expert Group on Artificial Intelligence of the European Commission: *A definition of AI: main capabilities and scientific disciplines* of 18 December 2018 and *Ethics Guidelines for Trustworthy AI Working Document for Stakeholders' Consultation* of 18 December 2018.

¹⁸ <https://ia.italia.it/assets/librobianco.pdf> (last visited 10 November 2018).

steps for a PA - AI ready.¹⁹ With particular reference to the Challenge in the “legal context”, the focus is on: transparency of administrative acts, legal responsibility, data protection, computer security and intellectual property.²⁰

With more attention to the issue of e-health, on the occasion of the conference “*Innovation at the service of health. Towards a Manifesto for an Intelligent Healthcare*” held on 23rd October 2017, organized by I-Com, Institute for Competitiveness, a Manifesto on AI in Healthcare was presented.²¹ The text derives from the activity of representatives of institutions, businesses, and civil society. It consists of eight goals to “*accompany the country in the challenge of digital health, and ensure that Italy can seize all the many benefits that come from the use of AI solutions*”. It is worth mentioning here: point 5 “*The role of the physician must remain central to the implementation of AI, no machine will replace it. The physician-patient relationship is fundamental*”; point 6 “*Privacy, security and personal data protection represent fixed points in any fields, especially in a sensitive area as healthcare*”; and, finally, point 7 “*It is of primary importance to create the skills of the future which, thanks be to the development of AI, will be able to offer great opportunities of employment*”.

The challenge in the health sector then becomes creating services, also based on the use of AI, which allow the health administration to increasingly empower the citizen/patient. This is true not only in the moment of real and perceived need, but above all throughout the previous phase that helps, through new forms of prevention, to reduce the use of health services by reducing the need for subsequent diagnosis and treatment.

3. Data protection regulation issues

The use of AI and, in general, of the algorithms that determine its action is obviously the focus of attention also in the field of data protection.²² We have already highlighted above how some declarations have already begun to delineate the critical aspects that need to be addressed when analysing application contexts like this one.²³ European lawmakers already established with Directive 95/46/EC the existence of a right not to be subject to a decision that produces legal effects or significantly af-

¹⁹ *White Paper on artificial intelligence at the service of the citizen, id.*, 2018, pp. 72-77.

²⁰ *ibidem*, pp. 55-56. For a recent interesting study on AI in the legal context, see U. PAGALLO, M. PALMIRANI, P. CASANOVAS, G. SARTOR, S. VILLATA (eds), *AI Approaches to the Complexity of Legal Systems*, AICOL 2015, AICOL 2016, AICOL 2016, AICOL 2017, AICOL 2017. Lecture Notes in Computer Science, vol. 10791, Springer, Cham, 2018.

²¹ https://www.i-com.it/wp-content/uploads/2017/10/manifesto_per_una_sanita_intelligente1.pdf (last visited 10 November 2018).

²² A point of reference for the analysis carried out in this paragraph is represented by G. NOTO LA DIEGA, *Against the Dehumanization of Decision-Making - Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information*, in *JIPITEC*, vol. 9, No. 3, 2018, in part. at pp. 16-24, available at <https://www.jipitec.eu/issues/jipitec-9-1-2018/4677> (last visited 26 November 2018). See also S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, Vol. 7, Issue 2, 2017, pp. 76-99 (also available at <https://ssrn.com/abstract=2903469> (last visited 26 November 2018)).

²³ The European Data Protection Supervisor in the “Recommendations on the EU’s options for data protection reform” (2015/C301/01) at para 3.1 writes: “*the lack of meaningful information about the algorithmic logic which develops these profiles and has an effect on the data subject*”.

fects the data subjects, in the event that such decision was based only on an automated processing of data that had the objective of evaluating certain personal aspects of the individual.²⁴ Furthermore, the right to know the logic involved in any automated processing of data was established.²⁵

Regulation 2016/679 “General Data Protection Regulation” (GDPR)²⁶ recently chimed in on the issue by clearly defining that the person involved should not be subject to decisions taken solely by “intelligent” virtual agents. In the event that this happens, and this decision has legal effects on the subject concerned, “*or similarly significantly affects him or her*”, he/she has the right to obtain human intervention, to express his/her point of view and to challenge, if necessary, that decision. On the other hand, the data controller is subject to the obligation to provide “*meaningful information about the logic involved*” in the decision taken by the AI. Responding to this regulatory provision presents significant difficulties for the data controller as often the same logic that governs the operation of AI appears cryptic or unknown (the so-called “black boxes” phenomenon).

On this issue Art. 29 Data Protection Working Party took a position with the “Guidelines on Automated individual decision-making and adoption of the objective of Regulation 2016/679” (WP251rev.01), adopted on 3 October 2017 and last revised and adopted on 6 February 2018 (hereinafter: “Art. 29 WP on Automated decision-making”). In this document, the importance of two concepts regarding the protection of personal data is emphasized: “accountability”, the key theme of the GDPR and the new compass of data controller action, and “transparency”, a principle that should underlie all the activities relating to the processing of personal data.

From a general perspective, the pivotal aspects of the new European discipline regarding the context analysed here can be summarized in the following pivotal points. First of all, the importance of some essential principles (established by Article 5 of the GDPR) has been emphasized: “transparency”, which is not only a fundamental principle of processing, but also a real right of the data subject, since the data controller is responsible for making known “*any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used*” (Recital 39 GDPR); “correctness”, strictly connected to the concept of clarity and transparency, with particular attention to the information provided to the data subject in order to make his/her adequately understanding not only the modalities of the processing but also its consequences. The “accountability” of the data controller (art. 24 GDPR) has been greatly increased, moving from a checklist logic with reference to the implementation of security measures and obligations to a case by case assessment with respect to the choices to be adopted and the demonstration of these choices (consider for example the other requirements as the record of processing activities pursuant to art. 30). Moreover, it has become essential to clarify and specify the legal basis that legitimizes the processing: the data controller is obliged to assess which is the most appropriate legal basis with respect to the processing that she intends to carry out; in this activity she is not free to

²⁴ Art. 15 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (now repealed by Regulation 2016/679 “GDPR”).

²⁵ Art. 12 letter a) and Recital 41 Directive 95/47/EC.

²⁶ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

choose the preferred legal basis, but must comply with the conditions set by GDPR with respect to the characteristics of each of the bases indicated in Articles 6 or 9 and always be able to demonstrate the correctness of the choice made. Furthermore, as already noted, the individual's right to oppose an automated processing pursuant to art. 22 GDPR has been explicitly stated. Finally, GDPR at art. 35 establishes that a "Data Protection Impact Assessment" (DPIA) is needed when there is <<a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person>> (para. 3, letter a)). It is worth stressing that this provision does not refer only to evaluations that are "solely" based on automated processing: the DPIA is necessary in case of any automated profiling run by means of artificial intelligence, machine learning or other technologies capable of producing effects on individuals, even if there is a human intervention in evaluating the findings of the machines. This requirement underlines once again the importance of the accountability concept within the new European Regulation.

In the following sub-sections, the aspects qualifying the rules concerning the processing of personal data in the specific context of AI will be dealt with in detail: the prohibition of automated decisions, the (automated) processing of special categories of data and information rights.

3.1 The prohibition of automated decisions

GDPR states that: "*the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her*" (art. 22).²⁷ The prohibition applies to decisions that are based "solely" on automated processing; the human oversight of the conclusion reached by the machine needs to be meaningful. Art. 29 Working Party recommends treating this right as a general prohibition.²⁸

The new European Regulation did not innovate significantly with respect to art. 15 of Directive 95/46/EC. But some distinctions need to be drawn. First of all, the only protection provided for the automated decision-making envisaged in the Directive was the possibility of expressing one's point of view; Article 22, para. 3, GDPR allows challenging the decision taken and exercising the right to obtain human intervention. Furthermore, GDPR explicitly includes consent as a legitimate basis that authorizes the automated decision-making processes (para. 2, letter c)). Finally, unlike the previous

²⁷ For further analysis on this general prohibition, see G. NOTO LA DIEGA, *Against the Dehumanization of Decision-Making*, *op. cit.*, pp. 17-19.

²⁸ Art. 29 WP on Automated decision-making, p. 12. The first interpretation in favour of a general prohibition of algorithmic decisions was provided by the French *Loi n° 78-17* of 6 January 1978 *relative à l'informatique, aux fichiers et aux libertés*, at art. 10. As for the contemporary legal framework, for example, Italian lawmakers established this prohibition in Art. 14 "Definition of profiles and the personality of the data subject". The Italian Privacy Code (legislative decree 30 June 2003, n. 196, as amended by the implementing legislative decree 10 August 2018, n. 101) no longer includes provisions in this regard as art. 22 GDPR is intended to be directly applicable.

provision of the law, it is no longer necessary for the data subject to request the contract in order to legitimize the automated decision-making process.²⁹

First of all, we may stress the fact that there is an express reference to profiling as an example of automated processing. The definition of “profiling” is provided by art. 4, pt. 4, GDPR: it “means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements”. Art. 29 Working Party has observed that “(a)utomated decisions can be made with or without profiling; profiling can take place without making automated decisions”.³⁰ This opinion clarified that a mere classification of an individual on the basis of known characteristics does not per se trigger profiling, but it will depend on the purpose of the classification.³¹ Therefore, the simple classification of individuals does not configure this concept, but rather the usage of data that the controller is willing to perform. This definition could continue to preserve aspects of indeterminacy since statistical data are always processed by companies to make business decisions, otherwise there would not be any reason to collect them.

Another keyword of this provision is represented by “legal effects”. The abovementioned right may be enforced only if the decision produces legal effects concerning one “or similarly significantly affects him or her” (art. 22, para. 1, GDPR).³² The notion of “legal effect” is quite straightforward, including all the scenarios where a decision affects a person’s rights based on laws or contracts. By this perspective, the term “similarly” may narrow the scope of the provision, and be seen as meaning that one does not have the right to object to algorithmic decision-making if the effect is not similar to a legal effect (e.g. significant distress or missed professional opportunities as a consequence of being permanently banned from a popular social network). Art. 29 Working Party’s guidelines indicate that “similarly” means that “the threshold for significance must be similar”.³³ Therefore, in order for a decision to fall within the scope of art. 22, it is sufficient that it profoundly affects the individual as much as a decision affecting his or her rights would. The concept is broad enough to encompass a vast number of scenarios, from e-recruiting to online behavioural advertising, as well as consumer manipulation.³⁴

²⁹ See S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, *op. cit.*, p. 13.

³⁰ Art. 29 WP on Automated decision-making, p. 8.

³¹ See the example provided by Art. 29 WP: *id.*, p. 7: “a business may wish to classify its customers according to their age or gender for statistical purposes and to acquire an aggregated overview of its clients without making any predictions or drawing any conclusion about an individual. In this case, the purpose is not assessing individual characteristics and is therefore not profiling”.

³² An author observes that: “This addition goes in the opposite direction to the one taken when the draft GDPR was first published and it had been suggested that art. 22 should cover not only decisions producing legal effects or which significantly affect data subjects, but also the “collection of data for the purpose of profiling and the creation of profiles as such””, in G. NOTO LA DIEGA, *Against the Dehumanization of Decision-Making*, *op. cit.*, p. 18.

³³ Art. 29 WP on Automated decision-making, p. 10.

³⁴ See G. NOTO LA DIEGA, *Some considerations on intelligent online behavioural advertising*, in *Revue du droit des technologies de l’information*, vol. 66-67, 2017, p. 53.

Several aspects of this discipline are not yet fully clear. For example, even the term “solely” may give rise to different interpretations about its real applicative scope. In the past it was quite easy to understand its meaning, since only a limited number of organizations adopted significant algorithmic decisions and the technologies used were so rudimentary that human intervention was necessary in order to review the data processed. The development of technologies based on AI has now made the border between completely automated processing and one in which human intervention plays a significant role somewhat opaque and elusive.

However, GDPR at art. 22, para. 2, provides some exceptions to this general prohibition on making automated decisions. The latter shall not apply if the decision is: a) necessary for entering into or for the performance of a contract between the data subject and a data controller;³⁵ b) authorized by the law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests (i.e. the case of fraud prevention or money laundering checks); c) based on the data subject's explicit consent. The applicability of these exceptions is not straightforward.

Automated decision-making that involves special categories of data (pursuant to art. 9, para. 1³⁶) is only allowed under certain conditions and if suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place (art. 22, para. 4). The first condition allowing the processing is addressed if the data subject has given explicit consent to the processing of those personal data for one or more specified purposes (art. 9, para. 2, letter a)). Another possibility is provided by a need for reasons of “substantial public interest”, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject (art. 9, para. 2, letter g)). With particular reference to the digital health context, the information relating to these specific categories of data also includes “data concerning health”, as defined in art. 4, pt. 15.³⁷ The main basis that may legitimize the processing, and deactivates the prohibition referred to in the first paragraph of art. 22, within this particular scenario appears then to be represented by explicit consent, with all the difficulties related to the information obligations connected to the processing which will be delved into in the following sections. How could an informed consent be obtained in relation to a process that is inherently non-transparent (a “black-box”)? Providing a completely positive answer to this question, on which we will try to add further information, is very difficult.

³⁵ According to the EU data protection authorities, the interpretation of “necessity” to enter into a contract has to be interpreted narrowly: “the controller must be able to show that this profiling is necessary, taking into account whether a less privacy-intrusive method could be adopted. If other effective and less intrusive means to achieve the same goal exist, then it would not be ‘necessary’”, Art. 29 WP on Automated decision-making, p. 23.

³⁶ Art. 9, para. 1, GDPR: “Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited”.

³⁷ Art. 4, pt. 15, GDPR: “‘data concerning health’ means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status”.

Finally, there should be measures in place to safeguard the data subject’s rights and freedoms and legitimate interests. Art. 22, para. 3 of the GDPR establishes that, in case of decisions needed for entering into or performing a contract (par. 2, let. a)) or based on the data subject’s explicit consent (para. 2, letter c)), “the data controller shall implement suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision”. Thus, this paragraph states that individuals will in any case have the right to obtain human intervention to express their point of view and to contest the decision. This provision would apply also to a possible telemedicine scenario. Where a “virtual coach” module is implemented, the application of this provision would lead to the consequence that even if the process of management of the information related to the pathology provided, directly or by means of smart items (glucometers, cardiographers, etc.), by the patient, and for the supply of directly applicable medical advice, was completely automated, the patient shall always have the possibility to obtain the intervention of a human operator (physician or nurses depending on the context) who has the necessary authority and knowledge to evaluate and, if needed, modify the decision made. The system should also, finally, allow the patient to express her point of view and possibly challenge the decision provided.³⁸

3.2 A right to explanation or only a right to be informed?

An “original sin” that afflicts AI, especially in contexts where the processing of personal data becomes pivotal, is linked to the fact that the *how* and the *why* of their functioning does not appear so clear or immediately understandable. Typically, these techniques cannot, indeed, explain the process that led them to reach the conclusion they did.³⁹ Not only: since they often base themselves on machine-learning methodologies, where more data are available, these data can be incorporated in order to improve future predictions, but also and above all to modify the same algorithms that make the system work. Therefore, the algorithms that are at the heart of these digital systems are not only opaque but also likely to change over time.⁴⁰

In these contexts, the expression “black boxes” is used to explain this phenomenon.⁴¹ The term, often quoted in computer science and engineering, indicates a system that, similarly to a “black box”, is essentially describable in its external aspects, i.e. only for how it reacts at output to a given input, but whose internal functioning is not visible or is unknown. Three different types of black boxes can be identified.⁴² The “organizational black box” emphasizes the fact that algorithms are often implemented by private, profit-maximizing entities that operate under minimal transparency obligations: therefore

³⁸ See Art. 29 WP on Automated decision-making, p. 27: “Human intervention is a key element. Any review must be carried out by someone who has the appropriate authority and capability to change the decision. The reviewer should undertake a thorough assessment of all the relevant data, including any additional information provided by the data subject”.

³⁹ See J. BURRELL, *How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms*, in *Big Data & Soc’y*, 2016, vol. 3, p. 1, 5.

⁴⁰ See W. NICHOLSON PRICE, *Artificial Intelligence in Health Care*, *op. cit.*, p. 10.

⁴¹ See F. PASQUALE, *Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge – Massachusetts, 2015.

⁴² See G. NOTO LA DIEGA, *Against the Dehumanisation of Decision-Making*, *op. cit.*, pp. 9-10.

there are no incentives or real obligations to clarify the way information is processed. Then, there are the “technical black boxes” linked to the fact that AI often makes the rationale of decisions intrinsically difficult to achieve (i.e. you may consider the so-called “neural networks” that are modelled on the human brain). Finally, the “legal black boxes” related to intellectual property rights that can have an impact on the management, and transparency, of these technologies.⁴³

This intrinsic lack of transparency has strong repercussions on the accountability issue, a principle that we have seen inspiring and characterizing the GDPR approach. Ensuring fair, lawful and transparent processing becomes, therefore, very difficult considering how much these systems operate machine-learning environments integrated into more general flows of operations, involving data coming from different sources, and with different degrees of reliability, human intervention and deployment of machine-learning products and services.⁴⁴

Transparency may be quoted as one of the founding principles of GDPR. The processing of personal data must be made as intelligible as possible for the data subject. The main instrument is the information notice. Directive 95/46/EC only provided for the right of access, which included the logic used in the decision algorithm (art. 12, letter a), Directive). With GDPR processing can be considered fair and transparent only if specific information is provided, which essentially concerns three aspects (articles 13 and 14 GDPR): the data controller must explain the existence of a decision-making algorithm; the data subject must be informed of the logic used; the algorithm must be opened in order to provide “meaningful information about (...) the significance and the envisaged consequences of such processing for the data subject” (art. 13, para. 2, letter f), GDPR).

The right to explanation can have different meanings, depending on the context to which it refers:⁴⁵ “system functionality”, which represents the logic, the meaning, the expected consequences and the general functionalities of an automatic decision-making system (e.g. system’s requirements specification, decision trees, criteria, etc.); “specific decisions”, which concerns the rationale, the reasons, the individual circumstances of a specific automated decision (e.g. weighting of features, machine-defined case-specific decision rules, etc.). Furthermore, we can refer to the moment in which the explanation intervenes: *ex ante* explanation that takes place before the automated decision-making process has been put in place (in this case only the system functionality can be addressed); *ex post* explanation that occurs after an automated decision-making process has been implemented (this can address both system functionality and the rationale of the specific decision).

A right to explanation can be founded on three legal bases.

It may derive from safeguards against automated decision-making, as set by art. 22, para. 3, GDPR: “the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision”. Some have critically observed that the right to explanation would not be explicitly mentioned in the articulation of GDPR but would

⁴³ *Id.*, pp. 11-16.

⁴⁴ See D. KAMARINOU ET AL., *Machine Learning with Personal Data*, in *Queen Mary School of Law Legal Studies Research Paper No. 247*, 2016, p. 22; C. KUNER ET AL., *Machine learning with personal data: is data protection law smart enough to meet the challenge?*, in *International Data Privacy Law*, 2017, vol. 7, issue 1, p. 1.

⁴⁵ See S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, *op. cit.*, p. 6-7.

find room only in the Recital 71: “In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision”.⁴⁶

The right in question can also be supported on the basis of the notification duties under Articles 13 and 14 commented upon by Recitals 60-62. In particular art. 13, para. 2, and art. 14, para. 2, state that the data controllers need to “provide the data subject with the following further information necessary to ensure fair and transparent processing”. Coming to our issue, according to art. 13, para. 2, letter f) and art. 14, para. 2, letter g), this information includes: “the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject”.

According to Article 29 Working Party the explanation of the logic involved would include details on the rationale behind, or the criteria relied on, in reaching the decision, without necessarily always attempting a complex explanation of the algorithms used or a disclosure of the full algorithm. The clarification on the level of details is important because otherwise individuals might understand the logic followed by the machine and act in such a way that they can take unfair advantages.⁴⁷ It is not possible to adopt a privacy information notice that would cover any type of machine learning or artificial intelligence technology. It should outline the main characteristics considered in reaching the decision, the source of this information and their relevance.

Finally, another reference to the right to explanation in GDPR may be pinpointed in the right of access established in art. 15 GDPR, which uses the same wording of art. 13, para. 2, letter f), and 14, para. 2, letter h) with reference to the fact that the data subject is allowed to use automatic decision-making and to obtain meaningful information about the significance, envisaged consequences, and logic involved.⁴⁸

As already mentioned, some scholars have criticized GDPR as it would entail a right to be informed, but not an explicit and specific right to explanation.⁴⁹ Other authors have instead emphasized the fact that articles 15 and 22 should be subject to will be open on public consultation extensive interpretation in order to be able to meet the need to give real meaning to the principle of transparency; they have therefore proposed a legibility stress test for the holder of the treatment.⁵⁰ We can hereby join

⁴⁶ *Id.*, pp. 7 ff. (the authors held that the omission of a right to explanation from Article 22 appears to be intentional, p. 11).

⁴⁷ See Art. 29 WP on Automated decision-making, p. 25.

⁴⁸ Watcher et al. claim that: “(...) as with notification duties in Articles 13-14, and regardless of when it is invoked by the data subject, the GDPR’s right of access only grants an explanation of automated decision making addressing system functionality, not the rationale and circumstances of specific decisions”. S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, *op. cit.*, p. 19.

⁴⁹ *Id.*, *passim*; L. EDWARDS, M. VEALE, *Slave to the algorithm? Why a ‘right to explanation’ is probably not the remedy you are looking for*, 16 *Duke L. & Tech. R.* 18 (2017).

⁵⁰ See G. MALGIERI, G. COMANDÉ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, 2017, vol. 7, issue 4, p. 243; J. POWLES, H. HODSON, *Google DeepMind and healthcare in an age of algorithms*, in *Health Technol.*, 2017, vol. 7, p. 351.

the position of other scholars stating that the rights of information provided by GDPR with reference to automated decision-making processes (as to the significance and consequences of the decision) may, indeed, be interpreted in such a way as to be considered a real right to explanation.⁵¹ Reasoning on the contrary could be dangerous and, above all, lead to a lowering of the guarantees recognized to the data subject, wasting the potential effect of the new European regulation in terms of protection of personal data.

4. Final remarks: algorithms, empathy and the importance of making mistakes

Medical practice has always used technology to improve the effectiveness and efficiency of its actions. Artificial intelligence represents, in order of time, the last chapter of this relationship. This technology seems to guarantee increasingly personalized medicine, and, hopefully frees from the possibility of making human mistakes.

Thus the adoption of AI in e-health promises several desirable results, but also hides possible dystopian scenarios. In this context, the law plays a fundamental role as it allows to govern in a “human” way a technology that could instead lead to phenomena of dehumanization.

Scholars such as Rodotà speak of “dictatorship of the algorithm”, evoking a phenomenon that is now global and not reserved only to the medical field. The question is linked to the relationship between collected data and decision, and, therefore it impacts on the relationship that is established between the person and the power of the information holders. The algorithm draws, indeed, the ways of functioning of large areas of our social organizations, and thus redistributes powers. It incarnates, from a certain perspective, the new forms of power and modifies its quality. When the algorithm becomes the very foundation of the power exercised by a specific individual or, rather, organization (see the case of Google) and all that concerns it is wrapped up in the utmost secrecy, then we are really facing the new version of the *arcana imperii*.⁵² This widespread awareness should lead us to adopt at least the “precautionary principle” and build an appropriate institutional framework, questioning whether and how the algorithm society can be truly democratic.⁵³

The issue has cultural and social implications too. The patient-doctor relationship traditionally relies on orality and is based on a purely human relationship. Such a relationship is based precisely on the cognitive ability to understand and feel those people's emotions in the specific situation that afflicts them. Often in the rhetoric that accompanies these new technologies, however, empathy is seen as a *bias* and therefore an argument in favour of decision systems exclusively based on non-human agents.⁵⁴ Like empathy, also “mistaking” plays a fundamental role. The successful television series *Westworld*⁵⁵ is about an amusement park populated by androids created to allow visitors a realistic and sometimes violent western-themed experience. Dr. Robert Ford, the creative director of the

⁵¹ See G. NOTO LA DIEGA, *Against the Dehumanization of Decision-Making*, op. cit., pp. 23-24.

⁵² The term “*arcana imperii*” literally means the “secrets of power” or the “principles of power” or “state”. It is found in two passages of the work of Tacitus, specifically in the *Historiae* (I, 4) and in the *Annales* (II, 36).

⁵³ S. RODOTÀ, *Il diritto di avere diritti*, Laterza, Bari, 2012, in part. 312-340.

⁵⁴ See G. NOTO LA DIEGA, *Against the Dehumanization of Decision-Making*, op. cit., p. 11.

⁵⁵ Famous US television series produced by Jonathan Nolan and Lisa Joy for HBO, based on the film *The World of Robots* (*Westworld*, 1973), written and directed by Michael Crichton.

park, played by the famous Anthony Hopkins, who continually updates the androids with the so-called “memories” to make them more and more like humans, at a certain point says: *“Mistakes’ is the word you’re too embarrassed to use. You ought not be. You’re the product of a trillion of them. Evolution forged the entirety of sentient life on this planet using only one tool: the mistake”*. It is, actually, needed to accept that these systems can be wrong and that they learn, together with us, by their errors.

AI must be informed by those “human, too human” aspects that characterize our actions in health care scenarios. It must be inspired by the deontology of the medical profession. Otherwise, if fed only by reports and cases, it will risk deviant diagnosis and therapies: splendidly rational, but not necessarily the best for the patient. AI represents a sort of mirror of ourselves and, at the same time, a possible tool to restore our ethics.⁵⁶

Acknowledgements

I would like to thank Chiara Di Natale, Rossana Ducato, Federica Giovanella, Rachele Zanchetta for their valuable feedback. All errors remain my own.

⁵⁶ In G. NOTO LA DIEGA, *Against the Dehumanization of Decision-Making*, op. cit., p. 33: *“In order to better understand how to make the human-algorithm cooperation work best, it has become crucial to shift the focus from the definition of algorithms, artificial intelligence etc. to the understanding of what makes us human”*.